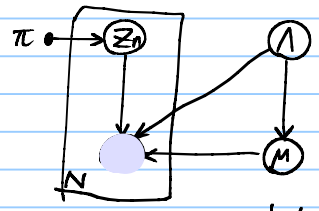
10.23



というモデルについて、確率変数の同時分布は

$$p(X,Z,\mu,\Lambda|\pi) = p(X|Z,\mu,\Lambda)\,p(Z|\pi)\,p(\mu|\Lambda)\,p(\Lambda) \quad \cdots ①$$

である。ここで潜在変数とパラメータの事後分布

$$p(Z,\mu,\Lambda|X,\pi)$$

の変分近似 $q(Z,\mu,\Lambda|\pi)$ を考える。

10.2章において π の分布を考えたとき変分近似を $q(Z,\pi,\mu,\Lambda) = q(Z)\,q(\pi,\mu,\Lambda)$ と分解できると仮定した

これは Z と π が独立と仮定してることを含んでいる。

そこで、ここの変分近似においても Z と μ,Λ は分解できるとし、また Z と π は独立であると仮定する。

$$q(Z,\mu,\Lambda|\pi) = q(Z|\pi)\,q(\mu,\Lambda) = q(Z)\,q(\mu,\Lambda) \quad \cdots ②$$

Z と μ,Λ は分解できると仮定　モデルより元々 μ,Λ は π とは独立なので　Z と π は独立と仮定

このとき変分下界は

$$\mathcal{L}(q(Z\mu\Lambda|\pi)) = \sum_Z \int q(Z\mu\Lambda|\pi) \ln \frac{p(X,Z,\mu,\Lambda|\pi)}{q(Z,\mu,\Lambda|\pi)}\, d\mu\, d\Lambda$$

$$= E[\ln p(X,Z,\mu,\Lambda|\pi)] - E[\ln q(Z,\mu,\Lambda|\pi)]$$

$$= E[\ln p(X|Z,\mu,\Lambda)\,p(Z|\pi)\,p(\mu|\Lambda)\,p(\Lambda)] - E[\ln q(Z)\,q(\mu,\Lambda)]$$

①　②

ここで

$$E[\ln p(X|Z,\mu,\Lambda)\,p(Z|\pi)\,p(\mu|\Lambda)\,p(\Lambda)]$$

$$= E[\ln p(Z|\pi)] + C$$

← π に依らない項

$E[\ln p(X|Z,\mu,\Lambda)] = \sum_Z \int q(Z)\,q(\mu,\Lambda) \ln p(X|Z,\mu,\Lambda)\, d\mu\, d\Lambda$

なので π に依らない

$E[\ln p(\mu|\Lambda)]$, $E[\ln p(\Lambda)]$ も同様

$E[\ln q(Z)]$, $E[\ln q(\mu,\Lambda)]$ も同様

Z と π は独立と仮定したので

$q(Z)$ は π に依らない

また

$$E[\ln q(Z)\,q(\mu,\Lambda)] = C$$

← π に依らない項

なので

$$\mathcal{L}(q(Z\mu\Lambda|\pi)) = E[\ln p(Z|\pi)] + C$$

← π に依らない項

$$= E\Big[\ln \prod_{n=1}^{N}\prod_{k=1}^{K} \pi_k^{z_{nk}}\Big] + C \quad (10.37)$$

$$= E\Big[\sum_{n=1}^{N}\sum_{k=1}^{K} z_{nk} \ln \pi_k\Big] + C$$

$$= \sum_{n=1}^{N}\sum_{k=1}^{K} \ln \pi_k\, E[z_{nk}] + C$$

$$= \sum_{n=1}^{N}\sum_{k=1}^{K} \ln \pi_k\, r_{nk} + C \quad (10.50)$$

となる。

π は下界 $\mathcal{L}(q)$ を最大化させるように決める。

π についての制約条件は

$$\sum_{k=1}^{K} \pi_k = 1$$

これよりラグランジアンは

$$L = \mathcal{L}(q(Z, \mu, \Lambda | \pi)) + \lambda \left(\sum_{k=1}^{K} \pi_k - 1 \right)$$

$$= \sum_{n=1}^{N} \sum_{k=1}^{K} \ln \pi_k r_{nk} + C + \lambda \left(\sum_{k=1}^{K} \pi_k - 1 \right)$$

停留条件は

$$\frac{\partial L}{\partial \pi_k} = 0$$

より

$$\sum_{n=1}^{N} \frac{1}{\pi_k} r_{nk} + \lambda = 0$$

$$\therefore \ \pi_k = \frac{-1}{\lambda} \sum_{n=1}^{N} r_{nk}$$

ここで

$$1 = \sum_{k=1}^{K} \pi_k = \sum_{k=1}^{K} \frac{-1}{\lambda} \sum_{n=1}^{N} r_{nk} = \frac{-1}{\lambda} \overbrace{\sum_{n=1}^{N} \overbrace{\sum_{k=1}^{K} r_{nk}}^{=1}}^{=N}$$

より

$$\lambda = -N$$

よって

$$\pi_k = \frac{1}{N} \sum_{n=1}^{N} r_{nk}$$

を得る。